DOCUMENT RESUME

ED 325 510 TM 015 748

AUTHOR Melancon, Janet G.; Thompson, Bruce
TITLE Measurement Characteristics of the MAA Math Placement Tests.
PUB DATE 1 Nov 90
NOTE 13p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (19th, New Orleans, LA, November 14-16, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Entrance Examinations; College Mathematics; Higher Education; *Item Analysis; Mathematics Achievement; *Mathematics Tests; *Screening Tests; *Student Placement; *Test Reliability; *Undergraduate Students
IDENTIFIERS *Placement Tests

ABSTRACT

This paper presents analyses of score reliability for a mathematics placement test developed by the Mathematical Association of America in 1984 for use with undergraduate college students. Subjects included all 589 students seeking admission to the college mathematics curricula at a private university in the South during the course of the study. Common misconceptions about reliability are also reviewed, such as the misconception that tests, as opposed to data, are reliable; and the idea that longer tests always yield more reliable data than do their shorter counterparts. The test was a revision of items previously reported. The revised version of the 50-item test included 2 subscales of 25 items each: an arithmetic scale, and an algebra scale. Reliability estimates and various item analysis statistics were reported. Revisions of the test yielded slightly better than expected alpha coefficients for the arithmetic and algebra subscales. Overall, the results suggest that the revised test can yield reasonably reliable data, and the test can be used to place students in an appropriate curriculum. One table contains the study data. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

jan90.wp 11/1/90

ED325510

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Bruce Thompson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

MEASUREMENT CHARACTERISTICS OF THE MAA MATH PLACEMENT TESTS

Janet G. Melancon
Loyola University

Bruce Thompson
Texas A&M University

Paper presented at the annual meeting of the Mid-South Educational Research Association, November 15, 1990.

TM015748

ERIC

ABSTRACT

The paper presents analyses of score reliability for math placement tests developed for use with undergraduate students. Subjects were all 589 students seeking admission to the math curricula during the course of the study. The paper also reviews some common misconceptions about reliability, e.g., the misconception that tests are reliable, and the misconception that longer tests always yield more reliable data than their shorter counterparts. It is found that the test and its subscales yield data with reasonably sound measurement integrity.

Counselors and faculty in mathematics departments have found the placement tests developed by the Mathematical Association of American (MAA) (1984) to be useful in providing optimal instruction for students. These tests weie developed for college students and can be used to place students in courses ranging from remedial programs through calculus. However, as noted in the user's guide, the measurement integrity of test results must be evaluated in each test application.

Too few researchers and educators realize that tests are not reliable or unreliable; rather, **data** have these characteristics, albeit data generated on a given measure administered with a given protocol to given subjects on given occasions. As Rowley (1976, p. 53) notes, "It needs to be established that an instrument itself is neither reliable nor unreliable." As Sax (1980, p. 261) explains,

> It is more accurate to talk about the reliability of measurements (data, scores, and observations) than the reliability of tests (questions, items, and other tasks). Tests cannot be stable or unstable, but observations can. Any reference to the "reliability of a test" should always be interpreted to mean the "reliability of measurements or observations [i.e., a particular set of data] derived from a test."

One important implication of the realization that reliability inures to data (rather than tests) is that reliability should generally be explored whenever data are collected.

Too few researchers realize that reliability is critical is to detecting substantive effects or in making placement decisions. As Locke, Spirduso and Silverman (1987, p. 28) note, "the correlation between scores from two tests cannot exceed the square root of the product for reliability in each test." Thus, if a researcher is correlating scores having a reliability of .9 with scores having a reliability of .6, the correlation cannot exceed .73. Prospectively, researchers must select measures that will allow detection of effects at the level desired; retrospectively, researchers must take reliability into account when interpreting findings.

The purpose of the present paper is to report two types of measurement studies conducted with one set of the MAA placement measures, i.e., a revised set of items employed in a previous report (Melancon & Thompson, 1989). First, various reliability analyses in which classical alpha coefficients were computed are reported; classical test theory is not without its limits (Algina, 1990; Eason, in press; Thompson, 1989; Webb, Rowley & Shavelson, 1988), but internal consistency reliability estimates can be useful in applied studies, such as the present study. Second, various item analysis statistics (Thompson & Levitov, 1985) are also reported, since these also bear upon reliability.

Subjects

Subjects in a previous study (Melancon & Thompson, 1989) of MAA placement test items were 539 undergraduate students at a

ERIC

private university in the south. The previous version of the 50-item test included three subscales, respectively measuring arithmetic (18 items), algebra (20 items), and finite math (12 items).

Subjects in the present study were 589 undergraduate students at the same university. The revised version of the 50-item test included two subscales of 25 items each, an arithmetic scale and an algebra scale. Subjects were all the students who were seeking admission into the undergraduate mathematics curriculum.

Coefficient alpha Analyses

Alpha coefficients for the arithmetic ($\underline{v}$=18), algebra ($\underline{v}$=20) and finite math ($\underline{v}$=12) subtest scores in the previous study (Melancon & Thompson, 1989) were .60, .82, and .47, respectively. Coefficient alpha for the total score based on 50 items was .83. Based on these analyses the finite math subscale was dropped, and additional items were added to the first two scales.

Application of the Spearman-Brown prophecy formula suggests that changing the arithmetic subscale a factor of $\underline{k}$=1.388888 (25 items / 18 items) would yield a new reliability of:

$$
\begin{aligned}
r' = & (k * r) / (1 + (k - 1) * r) \\
& (1.38 * 0.6) / (1 + (1.38 - 1) * 0.6) \\
& (0.83) / (1 + (0.38) * 0.6) \\
& (0.83) / (1 + 0.233) \\
& (0.83) / (1.233) \\
& 0.675,
\end{aligned}
$$

if the added items were of exactly the same quality as the previous items.

Similarly, changing the algebra subscale by a factor of $\underline{k}$=1.25

was expected to yield a new reliability of:

```
r' = ( k    *  r   )  /  ( 1 + ( k    - 1 ) *  r    )
     (1.25 *0.82 )  /  ( 1 + (1.25 - 1 ) * 0.82 )
     (       1.02 )  /  ( 1 + (0.25     ) * 0.82 )
     (       1.02 )  /  ( 1 +               0.205 )
     (       1.02 )  /  (                   1.205 )
                                            0.850
```

The revised subscales actually had alpha coefficients of .75 and .86, respectively. Alpha for the total scale was .89. The standard deviation for the subscales, respectively 3.9 and 5.7, appeared to be the basis for the obtained reliability estimate for the arithmetic scale data (.75), as against poor item quality. The students did better on arithmetic items, as might be expected, and were relatively homogeneous with respect to their performance in this skill area.

Variance of total scores is an important influence on calculated alpha, as can be seen from examination of the formula for computing alpha for dichotomously scored items:

$$(v / (v - 1)) * (1 - (S_{pq} / SD**2)),$$

where "S_{pq}" is the sum of the item variances, and *SD* is the standard deviation of the total test or scale scores. Since the variance of a dichotomously scored item equals the proportion of people who get the item right (p) times the proportion who get the item wrong (q = 1-p), even when item variance is at its maximum (p = q = .5), item variance tends to be a small value (p * q = .5 * .5 = .25) that has a limited influence on alpha. Thus, alpha is heavily influenced by the variance of the total scores on a scale or a test, and is larger as the score variance gets larger.

ERIC

It is a misconception that longer tests inherently yield more reliable data than shorter tests (Melancon & Thompson, 1990; Thompson, 1990). It is total variance (SD**2) that largely drives reliability estimates. Total variances equals the sum of the item variances (S_{pq}) plus 2.0 times the sum of (each unique interitem correlation coefficient times the product of the standard deviations of the two items in each unique pairwise combination of items). Thus, when new items are added to a test, if they are negatively correlated with the previous items, the total test variance will actually go down!

Item Analyses

Table 1 presents item difficulty (p) and SD ((p * q)**.5) statistics for the present study. The table also presents scale (v=25) alpha-if-deleted statistics for each item, and the corrected correlation between scores on each item ("0" or "1") and total scores on each scale (conceivably ranging "0" to "24") when the given item is omitted from the scale score (rIxS). Finally, the table presents test (v=50) alpha-if-deleted statistics for each item, and the corrected correlation between scores on each item ("0" or "1") and total scores on the test (conceivably ranging "0" to "49") when the given item is omitted from the scale score (rIxT).

INSERT TABLE 1 ABOUT HERE.

Discussion

The revisions of the MAA tests yielded slightly better than expected alpha coefficients for both the arithmetic (.75) and the algebra (.86) scales. The somewhat lower reliability of the arithmetic scale appears to be related to a higher mean on the scale and more homogeneous performance (SD=3.9), an occurrence that might be expected in a college setting. As indicated by the Table 1 results, the 50 items were all positively correlated with scale and test scores, thus each item adds to total score variance, and makes a positive contribution to score reliability.

Overall, the results suggest that the revised test can yield data that are reasonably reliable. No items detracting from test performance were noted. It is therefore suggested that the revised test may be useful in placing students into an appropriate curriculum, though the measurement integrity of data from the test should be continuously reviewed as part of future applications of the measure.

References

Algina, J. (1990). Elements of classical reliability theory and generalizability theory. In B. Thompson (Ed.), Advances in social science methodology (Vol. 1, pp. 137-169). Greenwich, CT: JAI Press.

Eason, S. (in press). Why generalizability theory yields better results than classical test theory: A primer with concrete examples. In B. Thompson (Ed.), Advances in educational research: Substantive findings, methodological developments (Vol. 1). Greenwich, CT: JAI Press.

Locke, L.F., Spirduso, W.W., & Silverman, S.J. (1987). Proposals that work: A guide for planning dissertations and grant proposals (2nd ed.). Newbury Park, CA: SAGE.

Mathematical Association of America. (1984). User's guide: The placement test program of the Mathematical Association of America (3rd ed.). Washington, DC: Author.

Melancon, J., & Thompson, B. (1989, January). Local norms and test characteristics for selected forms of the MAA Math Placement Tests. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston. (ERIC Document Reproduction Service No. ED 303 488)

Melancon, J., & Thompson, B. (1990). Maximizing test reliability by stepwise variable deletion: A case study with the Finding Embedded Figures Test. Perceptual and Motor Skills, 70, 99-110.

Rowley, G.L. (1976). The reliability of observational measures. American Educational Research Journal, 13, 51-59.

ERIC

Sax, G. (1980). Principles of educational and psychological measurement and evaluation (2nd ed.). Belmont, CA: Wadsworth.

Thompson, B. (1989, January). Why generalizability coefficients are an essential aspect of reliability assessment. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston.

Thompson, B. (1990). ALPHAMAX: A program that maximizes coefficient alpha by selective item deletion. Educational and Psychological Measurement, 50, 585-589.

Thompson, B., & Levitov, J.E. (1985). Using microcomputers to score and evaluate test items. Collegiate Microcomputer, 3, 163-168.

Webb, N.M., Rowley, G.L., & Shavelson, R.J. (1988). Using generalizability theory in counseling and development. Measurement and Evaluation in Counseling and Development, 21, 81-90.

Table 1
Item Analysis for the 50 Item Test

Item	p	SD	Scale alpha if Delete	rIxS	Test alpha if Delete	rIxT
Arithmetic Scale						
1	.90	.30	.75	.14	.89	.09
2	.94	.24	.75	.22	.89	.24
3	.96	.19	.76	.02	.89	.03
4	.82	.39	.74	.33	.89	.39
5	.77	.42	.73	.44	.89	.44
6	.91	.29	.75	.15	.89	.18
7	.77	.42	.75	.28	.89	.29
8	.62	.49	.73	.42	.89	.43
9	.86	.35	.74	.30	.89	.30
10	.90	.31	.75	.28	.89	.33
11	.65	.48	.74	.29	.89	.35
12	.96	.19	.75	.12	.89	.15
13	.87	.34	.74	.33	.89	.40
14	.57	.50	.76	.15	.89	.12
15	.55	.50	.75	.25	.89	.25
16	.63	.48	.73	.52	.89	.48
17	.50	.50	.74	.40	.89	.39
18	.34	.47	.75	.28	.89	.29
19	.66	.47	.75	.27	.89	.22
20	.83	.38	.74	.33	.89	.31
21	.88	.33	.74	.34	.89	.40
22	.71	.46	.74	.40	.89	.40
23	.45	.50	.75	.25	.89	.23
24	.77	.42	.74	.32	.89	.28
25	.84	.37	.75	.15	.89	.18
Algebra Scale						
26	.84	.36	.86	.41	.89	.43
27	.51	.50	.86	.50	.89	.49
28	.72	.45	.86	.53	.89	.52
29	.76	.43	.86	.53	.89	.54
30	.92	.28	.86	.33	.89	.36
31	.75	.43	.86	.50	.89	.52
32	.68	.47	.86	.48	.89	.49
33	.88	.33	.86	.37	.89	.38
34	.81	.39	.86	.27	.89	.31
35	.60	.49	.86	.53	.89	.48
36	.64	.48	.86	.33	.89	.33
37	.55	.50	.85	.60	.89	.58
38	.75	.43	.86	.33	.89	.32
39	.48	.50	.86	.42	.89	.40
40	.40	.49	.86	.50	.89	.47
41	.65	.48	.86	.49	.89	.51
42	.76	.43	.86	.44	.89	.42
43	.27	.44	.86	.50	.89	.47
44	.56	.50	.86	.53	.89	.49

45	.49	.50	.86	.36	.89	.33
46	.50	.50	.86	.32	.89	.30
47	.44	.50	.86	.36	.89	.35
48	.50	.50	.87	.20	.89	.21
49	.54	.50	.86	.25	.89	.28
50	.47	.50	.86	.42	.89	.45